Available at
POWERED BY SCIENCE @ DIRECT®
ELSEVIER
Differential Geometry and its Applications 20 (2004) 209–224
DIFFERENTIAL GEOMETRY AND ITS APPLICATIONS
www.elsevier.com/locate/difgeo

Gradient flows and double bracket equations

Tin-Yau Tam

Department of Mathematics, Auburn University, AL 36849-5310, USA

Received 29 September 2000

Communicated by T. Ratiu

Abstract

A unified extension of the gradient flows and the double bracket equations of Chu–Driessel and Brockett is obtained in the frame work of reductive Lie groups. We examine the gradient flows on the orbit in the Cartan subspace of a reductive Lie algebra, under the adjoint action. The results of Chu–Driessel and Brockett are corresponding to the reductive groups $GL(n,\mathbb{R})$ and $O(p,q)$.

MSC: primary 58F25; 65K10

Keywords: Reductive groups; Gradient flows; Orthogonal projections; Double bracket equations; Least squares optimizations

1. Introduction

Given a real $n \times n$ symmetric matrix x_0, the *isospectral surface* [7] of x_0 is defined to be the set $O(x_0) = \{kx_0k^{-1}\colon k \in O(n)\}$ where $O(n)$ is the orthogonal group. In other words, $O(x_0)$ is the set of $n \times n$ symmetric matrices cospectral with x_0, by the spectral theorem for real symmetric matrices. Chu and Driessel [7] studied the first two and Chu [6] considered the last one of the following optimization problems.

1. Given a real $n \times n$ symmetric matrix z, find $x \in O(x_0)$ that minimizes (maximizes)

$$-\frac{1}{2}\|x - z\|^2.$$

E-mail address: tamtiny@auburn.edu (T.-Y. Tam).

0926-2245/$ – see front matter
doi:10.1016/j.difgeo.2003.10.008

2. Find $x \in O(x_0)$ that minimizes (maximizes)
$$-\frac{1}{2}\|x - \operatorname{diag} x\|^2.$$
3. Given a diagonal $n \times n$ real matrix a, find $x \in O(x_0)$ that minimizes (maximizes)
$$-\frac{1}{2}\| \operatorname{diag} x - a\|^2,$$

where $\|\cdot\|$ is the Frobenius matrix norm, that is, $\|x\|^2 = \operatorname{tr} xx^{\mathrm{T}}$. Similar consideration was given [7] for $p \times q$ real matrices with prescribed singular values.

Brockett [3] independently studied problem 1 and later studied problems 1 and 3 in the context of compact connected semisimple Lie groups. The main results in [7] and [3] for problem 1 are the gradient flows of the functions
$$\Phi : O(n) \to \mathbb{R}, \quad k \mapsto \frac{1}{2}\left\|kx_0k^{-1} - z\right\|^2,$$
and
$$\varphi : O(x_0) \to \mathbb{R}, \quad x \mapsto \frac{1}{2}\|x - z\|^2.$$
The gradient flow of φ with respect to the normal metric (see the next section for the definition) on the tangent space of $O(x_0)$ at a point $x \in O(x_0)$ is
$$\frac{dx}{dt} = \bigl[x, [x, z]\bigr],$$
where $[\cdot,\cdot]$ denotes the Lie bracket. It turns out that this flow is indeed the projection of the gradient flow of Φ (with respect to the left-invariant Riemannian metric on the tangent space of $O(n)$):
$$\frac{dk}{dt} = k(t)\bigl[k(t)^{-1}x_0k(t), z\bigr],$$
along the curve $x(t) = k(t)^{-1}x_0k(t)$ lying in $O(x_0)$.

However, the results [7] for $p \times q$ matrices look different, apparently:
$$\frac{dX}{dt} = ZX^{\mathrm{T}}X - XZ^{\mathrm{T}}X + XX^{\mathrm{T}}Z - XZ^{\mathrm{T}}X.$$
It is the gradient flow of a similarly defined φ on the set of $p \times q$ matrices with prescribed singular values. This does not look like a double bracket equation but we will see that it is merely deceptive and the equation comes from a double bracket equation and the key is the reductive group $O(p, q)$.

The tool used in [7] and [3] is matrix differentiation. Chu and Driessel [7] extended the domain $O(n)$ of Φ to the set of $n \times n$ real matrices. By taking Fréchet derivative of the extended function, they obtained the gradient and then project it back to the tangent space of $O(n)$. Thus they obtained the double bracket equation and called the approach *projected gradient method* which is also used in some other papers of Chu [5,6]. The term “projected gradient” in [7] is different from the projected gradient flow with respect to the normal metric that we already mentioned. Our approach will be more direct.

One may rewrite
$$-\frac{1}{2}\|x - z\|^2 = B(x, z) + c,$$

where $c = -\frac{1}{2}(B(x,x) + B(z,z))$, a constant on $O(x_0)$; here $B(x,z) = \operatorname{tr} xz$ is the Killing form (up to a scalar multiple). Then the study is reduced to that of

$$\Phi : O(n) \to \mathbb{R}, \quad k \mapsto B\big(kx_0k^{-1}, z\big)$$

and

$$\varphi : O(x_0) \to \mathbb{R}, \quad x \mapsto B(x,z).$$

Bloch, Brockett and Ratiu [2] considered problem 1 in the context of compact Lie groups. Let K be a compact Lie group and denote by $B(\cdot,\cdot)$ the Killing form on $\mathfrak{k}$, the Lie algebra of K. They investigated the functions

$$\Phi : K \to \mathbb{R}, \quad \Phi(k) = B\big(\mathrm{Ad}\big(k^{-1}\big)x_0, z\big),$$

where $\mathrm{Ad} : K \to \mathrm{Aut}(\mathfrak{k})$ is the adjoint representation of K, and

$$\varphi : \Omega \to \mathbb{R}, \quad \varphi(x) = B(x,z),$$

where $\Omega = \mathrm{Ad}(K)x_0$ denotes the orbit of x_0 under the adjoint action of K. It is now known [2, Proposition 1.1] that the gradient flow of Φ relative to the left-invariant Riemannian metric on $\mathfrak{k}$ (whose value at the identity is minus the Killing form) is

$$\frac{dk}{dt} = (dL_k)\big[\mathrm{Ad}\big(k^{-1}\big)x_0, z\big],$$

where $L_k : K \to K$ is the left translation defined by $k' \mapsto kk'$, $k' \in K$. The projection of the flow onto Ω, obtained by setting $x(t) = \mathrm{Ad}(k(t)^{-1})x_0$ is then ([2, Corollary 1.2] and [4, Theorem 1 and Example 1])

$$\frac{dx}{dt} = \big[x, [x,z]\big],$$

which is known as the double bracket equation. However, it does not cover the results of Brockett–Chu–Driessel mentioned earlier. The adjoint orbit Ω may be identified with the corresponding coadjoint orbit. It is well known that coadjoint orbits in $\mathfrak{k}^*$ are symplectic manifolds and hence are even dimensional. But it is not the case in Brockett–Chu–Driessel's consideration, for example, $O(x_0)$ (the orbit of x_0 under the adjoint action of $SO(2)$) is 1-dimensional if x_0 is a generic 2×2 real symmetric matrix.

Prior to the works of Brockett and Chu–Driessel, Duistermaat, Kolk and Varadarajan had published a comprehensive paper [8] on the function Φ (see below) and their framework is a real connected semisimple Lie group G with finite center. Let $\mathfrak{a} = \mathfrak{k} + \mathfrak{p}$ be the Cartan decomposition of the Lie algebra $\mathfrak{g}$ of G. Let K be a maximal compact subgroup of G. They considered the smooth function

$$\Phi : K \to \mathbb{R}, \quad k \mapsto B\big(\mathrm{Ad}\big(k^{-1}\big)x_0, z\big),$$

where $x_0, z \in \mathfrak{p}$ are given. Indeed, the study of Φ goes back to Hermann and Takeuchi–Kobayashi [15, p. 214]. Bott [1] and Hunt (for compact Lie groups) [11] have used Φ. Critical points and the corresponding Hessians of Φ have been examined and the study implies some of the results in [7] with respect to problem 1.

The setting in [8] is more general than that in [2] if K in the consideration of [2] is connected. For if K is compact connected, then $K = ZK_s$ a commuting product of the center Z of K and the analytic subgroup K_s (semisimple) of $\mathfrak{k}_1 = [\mathfrak{k},\mathfrak{k}]$, where $\mathfrak{k} = \mathfrak{k}_1 + \mathfrak{z}$ [12, Theorem 4.29]. Now the adjoint actions of K and Z are trivial on $\mathfrak{z}$ and $\mathfrak{k}$ respectively. Thus $\mathrm{Ad}(K)x_0 = \mathrm{Ad}(K_s)\hat{x}_0 + \bar{x}_0$ where $x_0 = \hat{x}_0 + \bar{x}_0$ with

$\hat{x}_0 \in \mathfrak{k}_1$ and $\bar{x}_0 \in \mathfrak{z}$. So we can assume that K is semisimple. Now if K is semisimple, it can be viewed as a maximal compact subgroup of its complexification $K_{\mathbb{C}}$ whose Lie algebra is just $\mathfrak{g} := \mathfrak{k} + i\mathfrak{k}$ where $i = (-1)^{1/2}$. Thus the actions of K on $\mathfrak{k}$ and $\mathfrak{p} = i\mathfrak{k}$ are isomorphic via multiplication by i [15, p. 214].

In Section 2 an extension of the results of Chu and Driessel and Brockett on problem 1 is obtained, namely the gradient flow equations, as well as on problems 2 and 3. The results also extend those of Brockett [4] in which the three problems are studied in the context of compact connected semisimple Lie groups.

In Section 3, a brief discussion on the global extrema of the three optimization problems is given.

2. Gradient flows and double bracket equations

A reductive group consists of 4-tuples (G, K, θ, B), where G is a Lie group, K is a compact subgroup of G, θ is a Lie algebra involution of the Lie algebra $\mathfrak{g}$ of G, and B is a nondegenerate, $\mathrm{Ad}(G)$-invariant, symmetric, bilinear form on $\mathfrak{g}$ such that

(1) $\mathfrak{g}$ is reductive (meaning, $\mathfrak{g} = \mathfrak{g}_1 + \mathfrak{z}$, where $\mathfrak{g}_1 = [\mathfrak{g}, \mathfrak{g}]$ is semisimple and $\mathfrak{z}$ is the center of $\mathfrak{g}$),
(2) $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ (called the *Cartan decomposition*), where $\mathfrak{k}$ is the Lie algebra of K and is the $+1$ eigenspace and $\mathfrak{p}$ is the -1 eigenspace under the action of θ,
(3) $\mathfrak{k}$ and $\mathfrak{p}$ are orthogonal with respect to B, and B is negative definite on $\mathfrak{k}$ and positive definite on $\mathfrak{p}$,
(4) the map $K \times \exp \mathfrak{p} \to G$ given by multiplication is a surjective diffeomorphism,
(5) for every $g \in G$, the automorphism $\mathrm{Ad}(g)$ of $\mathfrak{g}$, extended to the complexification $\mathfrak{g}^{\mathbb{C}}$ of $\mathfrak{g}$ is contained in $\mathrm{Int}\, \mathfrak{g}^{\mathbb{C}}$, and

sometimes G is simply called a *reductive Lie group* [12, p. 384]. Now K is a maximal compact subgroup of G [12, Proposition 7.19].

Example 2.1.

(1) Let G be a (connected) semisimple Lie group with finite center, let B be the Killing form on $\mathfrak{g}$ which is the Lie algebra of G, let $\theta : \mathfrak{g} \to \mathfrak{g}$ be a Lie algebra involution such that the form $B_\theta(x, y) := -B(x, \theta(y))$ is positive definite (called a *Cartan involution*), let $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$, and let K be the analytic subgroup of G with Lie algebra $\mathfrak{k}$. Then G is reductive.
(2) Let G be a compact Lie group satisfying property (5). Then $K = G$, $\theta = \mathrm{id}$ and B is the negative of a $\mathrm{Ad}(G)$-invariant inner product.

Since B is $\mathrm{Ad}(G)$-invariant, it is $\mathrm{ad}(\mathfrak{g})$-invariant as well, which means

$$B\big([x, y], z\big) = -B\big(y, [x, z]\big), \quad x, y, z \in \mathfrak{g}.$$

Since θ is a Lie algebra involution, that is, an automorphism of the Lie algebra $\mathfrak{g}$ with $\theta^2 = \mathrm{id}$, condition (2) implies

$$[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{k}, \qquad [\mathfrak{k}, \mathfrak{p}] \subset \mathfrak{p}, \qquad [\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}.$$

Indeed, θ leaves $\mathfrak{z}$ and $\mathfrak{g}_1$ stable and the restriction of θ to $\mathfrak{g}_1$ is a Cartan involution [12, Proposition 7.19].

Throughout our discussion, we assume that G is reductive and denote by Ω an orbit in $\mathfrak{p}$ under the adjoint action of K, that is, $\Omega := \mathrm{Ad}(K)x$ for some $x \in \mathfrak{p}$. Given $x \in \mathfrak{p}$, denote by $\mathfrak{k}_x$ the centralizer of x in $\mathfrak{k}$, that is, $\mathfrak{k}_x := \{y \in \mathfrak{k}\colon [x, y] = 0\}$. In other words $\mathfrak{k}_x$ is the intersection of centralizer of x in $\mathfrak{g}$, that is, $\mathfrak{g}_x := \{y \in \mathfrak{g}\colon \mathrm{ad}(y)x = 0\}$ and $\mathfrak{k}$. Let $T_x\Omega$ be the tangent space of Ω at the point x.

Lemma 2.2.

(1) $T_x\Omega = \mathrm{ad}(\mathfrak{k})x = \mathrm{ad}(x)\mathfrak{k}$.
(2) *The restriction of* ad *on* $\mathfrak{k}_x^{\perp}$ *(the orthogonal complement of* $\mathfrak{k}_x$ *in* $\mathfrak{k}$ *with respect to* $B(\cdot,\cdot)$*),* $a_x := \mathrm{ad}(x)|_{\mathfrak{k}_x^{\perp}} : \mathfrak{k}_x^{\perp} \to T_x\Omega$*, is a vector space isomorphism.*
(3) $(T_x\Omega)^{\perp} = \mathfrak{p} \cap \mathfrak{g}_x$ *where* $(T_x\Omega)^{\perp}$ *denotes the orthogonal complement of* $T_x\Omega$ *in* $\mathfrak{p}$ *and* $\mathfrak{g}_x = \{y \in \mathfrak{g}\colon [x, y] = 0\}$ *is the centralizer of* x *in* $\mathfrak{g}$.

Proof. For any $y \in \mathfrak{k}$, $\frac{d}{dt}|_{t=0}\,\mathrm{Ad}(e^{ty})x = [y, x] = \mathrm{ad}(y)x$ and thus (1) follows. Now the kernel of $\mathrm{ad}(x)|_{\mathfrak{k}}$ is $\mathfrak{k}_x$ so (2) follows immediately. For $p \in \mathfrak{p}$ and $k \in \mathfrak{k}$, the relation $B(p, [x, k]) = B([p, x], k)$ together with $[p, x] \in \mathfrak{k}$ implies (3). □

Decompose orthogonally, relative the $B(\cdot,\cdot)$, $\mathfrak{k} = \mathfrak{k}_x + \mathfrak{k}_x^{\perp}$. Let $\mathrm{ad}(h)x, \mathrm{ad}(k)x \in T_z\Omega$ where $h, k \in \mathfrak{k}_x^{\perp}$. The inner product $\langle \mathrm{ad}(h)x, \mathrm{ad}(k)x\rangle_N := -B(h, k)$ is sometimes called the *normal metric* on $T_x\Omega$. In other words, $\langle u, v\rangle_N = -B(a_x^{-1}(u), a_x^{-1}(v))$ where $u, v \in T_x\Omega$.

Let $\varphi : \Omega \to \mathbb{R}$ be a smooth function. For $x \in \Omega$, we denote by $(\nabla\varphi)_x \in T_x\Omega = \mathrm{ad}(\mathfrak{k}^{\perp})x \subset \mathfrak{p}$ (since $[\mathfrak{k}, \mathfrak{p}] \subset \mathfrak{p}$) the gradient of φ with respect to $B(\cdot,\cdot)$, that is, $(d\varphi)_x(\cdot) = B((\nabla\varphi)_x, \cdot)$.

The following is an extension of [3, p. 82], [7, p. 1055] and [4, Theorem 1].

Theorem 2.3. *If* $\varphi : \Omega \to \mathbb{R}$ *is a smooth function, then the corresponding gradient flow with respect to the normal metric on* Ω *is given by*

$$\frac{dx}{dt} = \bigl[x, [x, (\nabla\varphi)_x]\bigr].$$

Along this flow

$$\frac{d\varphi}{dt} = -B\bigl(\bigl[x, (\nabla\varphi)_x\bigr], \bigl[x, (\nabla\varphi)_x\bigr]\bigr).$$

As $t \to \infty$*,* x *approaches an equilibrium point* $x(\infty) \in \Omega$ *and* φ *approaches a constant.*

Proof. By definition, the gradient of φ with respect to the normal metric, at a point $x \in \Omega$, denoted by $(\nabla_N\varphi)_x$, is given by the formula

$$(d\varphi)_x u = \bigl\langle(\nabla_N\varphi)_x, u\bigr\rangle_N, \quad \text{for all } u \in T_x\Omega,$$

that is, $B((\nabla\varphi)_x, u) = -B(a_x^{-1}((\nabla\varphi)_x), a_x^{-1}(u))$. Let $u = [x, y]$ where $y \in \mathfrak{t}_x^{\perp}$. Notice that

$$-B\bigl(a_x^{-1}\bigl(\bigl[x, \bigl[x, (\nabla\varphi)_x\bigr]\bigr]\bigr), a_x^{-1}(u)\bigr) = -B\bigl(\bigl[x, (\nabla\varphi)_x\bigr], y\bigr) = B\bigl((\nabla\varphi)_x, [x, y]\bigr) = B\bigl((\nabla\varphi)_x, u\bigr).$$

Thus

$$(\nabla_N\varphi)_x = \bigl[x, \bigl[x, (\nabla\varphi)_x\bigr]\bigr].$$

Along the flow given by $\frac{dx}{dt} = [x, [x, (\nabla\varphi)_x]]$,

$$\frac{d\varphi}{dt} = (d\varphi)_x\left(\frac{dx}{dt}\right) = B\big((\nabla\varphi)_x, \big[x, \big[x, (\nabla\varphi)_x\big]\big]\big) = -B\big(\big[x, (\nabla\varphi)_x\big], \big[x, (\nabla\varphi)_x\big]\big) \geqslant 0,$$

since $B(\cdot, \cdot)$ is negative definite on $\mathfrak{k}$ and $[\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}$. So φ is increasing along the gradient flow. Since K is compact, so is the orbit Ω. Thus φ approaches to a constant. It follows that $[x, (\nabla\varphi)_x] \to 0$ and thus $\frac{dx}{dt} \to 0$, that is, x approaches an equilibrium point in Ω. □

Remark. The double bracket equation has a geometric meaning. For a symmetric space $\mathfrak{p}$, the curvature tensor [9, p. 215] at the origin is $R_0(x, y)z = -[[x, y], z]$ where $x, y, z \in \mathfrak{p}$. Thus the gradient flow takes the form

$$\frac{dx}{dt} = R_0\big(x, (\nabla\varphi)_x\big)x.$$

Remark. For the sake of computation, sometimes, it would be easier to get $\varphi_x = (\nabla\varphi)_x + \hat{\varphi}_x \in \mathfrak{p}$, where $\hat{\varphi}_x \in (T_x\Omega)^{\perp}$, the orthogonal complement of $T_x\Omega$ in $\mathfrak{p}$. Of course

$$B(\varphi_x, u) = B\big((\nabla\varphi)_x + \hat{\varphi}_x, u\big) = B\big((\nabla\varphi)_x, u\big) \quad \text{for all } u \in T_x\Omega.$$

By Lemma 2.2,

$$[x, \varphi_x] = \big[x, (\nabla\varphi)_x\big],$$

and thus the gradient flow can be expressed as

$$\frac{dx}{dt} = \big[x, [x, \varphi_x]\big].$$

Example 2.4. Let $\varphi : \Omega \to \mathbb{R}$ be given by

$$\varphi(x) = B(x, z), \quad z \in \Omega.$$

Since $B(\cdot, \cdot)$ is Ad-invariant, we may assume that $z \in \mathfrak{a}$ where $\mathfrak{a}$ is a maximal abelian subalgebra in $\mathfrak{p}$. Indeed we may assume that $z \in \mathfrak{a}_+$ which is a fixed (closed) fundamental Weyl chamber of $\mathfrak{a}$ [10, Lemma 3.4]. Since elements of $T_x\Omega$ are of the form $\mathrm{ad}(y)x$, $y \in \mathfrak{k}_x^{\perp}$ by Lemma 2.2, and

$$\left.\frac{d}{dt}\right|_{t=0} \mathrm{Ad}\big(e^{ty}\big)x = \mathrm{ad}(y)x,$$

we have

$$(d\varphi)_x\big(\mathrm{ad}(y)x\big) = \left.\frac{d}{dt}\right|_{t=0} B\big(\mathrm{Ad}\big(e^{ty}\big)x, z\big) = B\big(\mathrm{ad}(y)x, z\big),$$

and thus $\varphi_x = z$. The gradient flow is then

$$\frac{dx}{dt} = \big[x, [x, z]\big].$$

Let $x(\infty) = \lim_{t\to\infty} x(t)$. Then $[x(\infty), z] = 0$. If $z \in \mathfrak{a}_+$ is regular, its centralizer in $\mathfrak{p}$ is $\mathfrak{a}$. Thus $x(\infty) \in \mathfrak{a}$. So this provides a means to find an element in $\Omega \cap \mathfrak{a}$. Though z is regular, $x(\infty) \in \Omega$ is not necessarily so.

We remark that the gradient flow is also valid for $\psi(x) := -\frac{1}{2}B(x-z, x-z)$ since $\psi(x) = \varphi(x) + c$ where $c = -\frac{1}{2}[B(x,x) + B(z,z)]$ is a constant. It is because that $z \in \mathfrak{p}$ is fixed and $B(\cdot,\cdot)$ is Ad-invariant and thus $B(x,x)$ remains constant on Ω.

(a) Chu–Driessel–Brockett's double bracket equation.

When $G = GL(n,\mathbb{R})$, $K = O(n)$, $B(x,y) = \operatorname{tr} xy$, $\theta : x + y \mapsto x - y$, $x \in \mathfrak{k}$ and $y \in \mathfrak{p}$, where $\mathfrak{p}$ is the set of real $n \times n$ symmetric matrices, $\mathfrak{k}$ is the set of real $n \times n$ skew symmetric matrices, $\mathfrak{a} \subset \mathfrak{p}$ is the set of real diagonal matrices. The gradient flow of φ or ψ becomes

$$\frac{dx}{dt} = \bigl[x, [x,z]\bigr],$$

which is given in [7, p. 1055] and [3, p. 82]. Finding an element in $\Omega \cap \mathfrak{a}$ amounts to the computation of the eigenvalues of a real $n \times n$ symmetric matrix x if we pick a diagonal symmetric matrix z with distinct eigenvalues, that is, regular element. Similar results hold for $GL(n,\mathbb{C})$, that is, the Hermitian case.

(b) Chu–Driessel's equation.

Now if $G = O(p,q)$ [16] ($p \leqslant q$ without loss of generality), $K = O(p) \times O(p)$, $B(x,y) = \operatorname{tr} xy$, $x, y \in \mathfrak{g}$.

$$\begin{aligned}
\mathfrak{o}_{p,q} &= \left\{ \begin{bmatrix} X_1 & Y \\ Y^{\mathrm{T}} & X_2 \end{bmatrix} :\ X_1^{\mathrm{T}} = -X_1,\ X_2^{\mathrm{T}} = X_2,\ Y \in \mathbb{R}_{p\times q} \right\},\\
K &= O(p) \times O(q),\\
\mathfrak{k} &= \mathfrak{o}(p) \oplus \mathfrak{o}(q),\\
\mathfrak{p} &= \left\{ \begin{bmatrix} 0 & Y \\ Y^{\mathrm{T}} & 0 \end{bmatrix} :\ Y \in \mathbb{R}_{p\times q} \right\},\\
\mathfrak{a} &= \bigoplus_{1 \leqslant j \leqslant q} \mathbb{R}(E_{j,p+j} + E_{p+j,j}),
\end{aligned}$$

where $E_{i,j}$ is the $(p+q) \times (p+q)$ matrix and 1 at the (i,j) position is the only nonzero entry. Now, on $\mathfrak{p}$,

$$B\left(\begin{bmatrix} 0 & X \\ X^{\mathrm{T}} & 0 \end{bmatrix}, \begin{bmatrix} 0 & Y \\ Y^{\mathrm{T}} & 0 \end{bmatrix} \right) = 2 \operatorname{tr} XY^{\mathrm{T}},$$

and the adjoint action of K on $\mathfrak{p}$ is given by

$$\begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} 0 & S \\ S^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 0 & USV^{\mathrm{T}} \\ VS^{\mathrm{T}}U^{\mathrm{T}} & 0 \end{bmatrix},$$

where $U \in O(p), V \in O(q)$. We identify $\mathfrak{p}$ with $\mathbb{R}_{p\times q}$ and thus $\mathfrak{a}$ is then identified with the set of real $p \times q$ "diagonal" matrices. We may choose $\mathfrak{a}_+ = \{\bigoplus_{1 \leqslant j \leqslant p} a_j E_{j,j}\colon a_1 \geqslant \cdots \geqslant a_p \geqslant 0\}$. Indeed, if $X \in \mathbb{R}_{p\times q}$, then the element in the singleton set $\operatorname{Ad}(K)X \cap \mathfrak{a}_+$ [10, Lemma 3.4] yields the vector of singular values of X, whose entries are arranged in nonincreasing order. The action of K on $\mathfrak{p}$ is then orthogonal equivalence, that is, $H \mapsto UHV^{\mathrm{T}}$ where $U \in O(p)$, $V \in O(q)$. By direct computation, the gradient flow

$$\frac{dx}{dt} = \bigl[x, [x,z]\bigr]$$

becomes [7, p. 1059]

$$\frac{dX}{dt} = ZX^{\mathrm{T}}X - XZ^{\mathrm{T}}X + XX^{\mathrm{T}}Z - XZ^{\mathrm{T}}X,$$

where

$$x(t) = \begin{bmatrix} 0 & X(t) \\ X(t)^{\mathrm{T}} & 0 \end{bmatrix}, \qquad z = \begin{bmatrix} 0 & Z \\ Z^{\mathrm{T}} & 0 \end{bmatrix}.$$

Now the flow is continuous on Ω. When $p = q$ and $x \in \Omega$ is nonsingular, Ω has two components, namely, Ω_+ (the matrices with positive determinant) and Ω_- (the matrices with negative determinant). So if one starts with x such that $\det x > 0$, say, then the flow remains on Ω_+. Of course, one can get that from the semisimple Lie group $SO(p,p)$ [9, p. 445] whose Lie algebra $\mathfrak{so}_{n,n}$ is the split form of real simple Lie algebra $\mathfrak{so}(2n,\mathbb{C})$.

We remark that the gradient flow holds for reductive $U(p,q)$ [9, p. 444], that is, the complex case, in which transpose becomes complex conjugate transpose.

The gradient flow can be interpreted as the projection of the gradient flow of

$$\Phi : K \to \mathbb{R}, \quad k \mapsto B\big(\mathrm{Ad}\big(k^{-1}\big)x_0, z\big),$$

with respect to the left-invariant Riemannian metric on $\mathfrak{k}$, where $x_0 \in \Omega$. The left translation $L_h : K \to K$ is the analytic diffeomorphism $k \mapsto hk$, $k \in K$ [9, p. 99]. The tangent space of K at the point k, denoted by $T_k K$, is $(dL)_k\mathfrak{k}$. Now the left-invariant Riemannian metric is simply $\langle p, q\rangle = -B((dL)_k^{-1}p, (dL)_k^{-1}q)$ where $p, q \in T_k K$. It is obvious that $\Phi = \varphi \circ p_{x_0}$ where $p_{x_0} : K \to \Omega$ is the natural projection $k \mapsto \mathrm{Ad}(k)x_0$, where $x_0 \in \Omega$.

Lemma 2.5. *Let $p_{x_0} : K \to \Omega$, $k \mapsto \mathrm{Ad}(k^{-1}(t))x_0$ be the projection with respect to a base point $x_0 \in \Omega$. Then*

$$(dp_{x_0})_k = \mathrm{ad}\big(\mathrm{Ad}\big(k^{-1}\big)x_0\big) \circ (dL)_k^{-1}.$$

Proof. Notice that for each $(dL)_k r \in T_k K$, where $r \in \mathfrak{k}$,

$$\begin{aligned}(dp_{x_0})_k\big((dL)_k r\big) &= \frac{d}{dt}\bigg|_{t=0} \mathrm{Ad}\big(ke^{tr}\big)^{-1}x_0 = \frac{d}{dt}\bigg|_{t=0} \mathrm{Ad}\big(e^{-tr}\big)\,\mathrm{Ad}\big(k^{-1}\big)x_0 \\ &= \big[\mathrm{Ad}\big(k^{-1}\big)x_0, r\big] = \big[\mathrm{Ad}\big(k^{-1}\big)x_0, (dL)_k^{-1}(dL)_k r\big].\end{aligned}$$

In other words, $(dp_{x_0})_k(\cdot) = [\mathrm{Ad}(k^{-1})x_0, (dL)_k^{-1}(\cdot)]$. $\square$

The following extends the results in [7, p. 1054] and [2, Proposition 1.1 and Corollary 1.2, 1.3] if K is connected.

Theorem 2.6. *The gradient flow of $\Phi : K \to \mathbb{R}$ defined by $\Phi(k) = B(\mathrm{Ad}(k^{-1})x_0, z)$, where $x_0 \in \Omega$, with respect to the left-invariant Riemannian metric on $\mathfrak{k}$ is*

$$\frac{dk}{dt} = (dL)_{k(t)}\big[\mathrm{Ad}\big(k(t)^{-1}\big)x_0, z\big].$$

The projection of the gradient flow onto the adjoint orbit Ω obtained by setting $x(t) = \mathrm{Ad}(k(t)^{-1})x_0$ is the double bracket equation

$$\frac{dx}{dt} = \big[x, [x, z]\big].$$

Proof. Each element of $T_k K$ is of the form $v = (dL)_k r$ where $r \in \mathfrak{k}$. By definition the gradient $(\nabla\Phi)_k$ of Φ at the point k, with respect to the left-invariant Riemannian metric, is given by

$$(d\Phi)_k v = \big\langle (\nabla\Phi)_k, v\big\rangle.$$

Now

$$\begin{aligned}(d\Phi)_k v &= \frac{d}{dt}\bigg|_{t=0} \Phi\big(ke^{tr}\big) = \frac{d}{dt}\bigg|_{t=0} B\big(\mathrm{Ad}\big(ke^{tr}\big)^{-1}x_0, z\big) = \frac{d}{dt}\bigg|_{t=0} B\big(\mathrm{Ad}\big(k^{-1}\big)x_0, \mathrm{Ad}\big(e^{tr}\big)z\big)\\ &= B\big(\mathrm{Ad}\big(k^{-1}\big)x_0, [r, z]\big) = B\big(\big[z, \mathrm{Ad}\big(k^{-1}\big)x_0\big], r\big) = \big\langle (dL)_k\big[\mathrm{Ad}\big(k^{-1}\big)x_0, z\big], v\big\rangle.\end{aligned}$$

Thus the gradient flow takes the desired form. By Lemma 2.5,

$$\frac{dx}{dt} = (dp_{x_0})_k\left(\frac{dk}{dt}\right) = \big[x, (dL)_k^{-1}(dL)_k[x, z]\big] = \big[x, [x, z]\big]. \qquad \square$$

Example 2.7. When $G = GL(n, \mathbb{R})$, the gradient flow ([7, p. 1054], [3, p. 81]) is

$$\frac{dk}{dt} = k\big[k^{-1}x_0k, z\big] = k\big(k^{-1}x_0kz - zk^{-1}x_0k\big).$$

When $G = O(p, q)$, let

$$k = \begin{bmatrix} U & 0\\ 0 & V\end{bmatrix} \in O(p) \times O(q), \qquad x = \begin{bmatrix} 0 & X\\ X^{\mathrm{T}} & 0\end{bmatrix} \in \Omega, \qquad z = \begin{bmatrix} 0 & Z\\ Z^{\mathrm{T}} & 0\end{bmatrix}.$$

Thus the gradient flow consists of a pair of equations [7, p. 1059]

$$\frac{dU}{dt} = U\big(U^{\mathrm{T}}XVZ^{\mathrm{T}} - ZV^{\mathrm{T}}X^{\mathrm{T}}U\big), \qquad \frac{dV}{dt} = V\big(V^{\mathrm{T}}X^{\mathrm{T}}UZ - Z^{\mathrm{T}}U^{\mathrm{T}}XV\big).$$

We remark that the gradient flow holds for reductive $U(p, q)$ (transpose becomes complex conjugate transpose) or the simple $SU(p, q)$. The underlying function is then

$$\varphi : \Omega \to \mathbb{R}, \quad \varphi(x) = B(x, z) = 2\,\mathrm{Re}\,\mathrm{tr}\, UXV^*Y,$$

where

$$x = \begin{bmatrix} 0 & X\\ X^* & 0\end{bmatrix} \in \Omega, \qquad z = \begin{bmatrix} 0 & Z\\ Z^* & 0\end{bmatrix}.$$

We remark that in [4, p. 81, Example 2], the function ψ should have Re in the definition.

The following extends a result in [7, p. 211].

Example 2.8. Let $\mathfrak{a}$ be a maximal abelian subalgebra in $\mathfrak{p}$. Let $\mathfrak{a}^{\perp}$ denote the orthogonal complement in $\mathfrak{p}$ of $\mathfrak{a}$ and let $\pi : \mathfrak{p} \to \mathfrak{a}$ denote the orthogonal projection of $\mathfrak{p}$ onto $\mathfrak{a}$. It is known that [10, Proposition 3.5]

$\mathfrak{a}^\perp = [\mathfrak{k}, \mathfrak{a}]$ which implies that π is independent of the choice of B. Given $a \in \mathfrak{a}$, let $\varphi : \Omega \to \mathbb{R}$ be

$$\varphi(x) = -\frac{1}{2}B\big(\pi(x) - a, \pi(x) - a\big), \quad x \in \Omega.$$

Similar to the previous example, for all $y \in \mathfrak{k}_x^\perp$, we have

$$\begin{aligned}(d\varphi)_x\big(\mathrm{ad}(y)x\big) &= -\frac{1}{2}\frac{d}{dt}\bigg|_{t=0} B\Big(\pi\big(\mathrm{Ad}\big(e^{ty}\big)x\big) - a, \pi\big(\mathrm{Ad}\big(e^{ty}\big)x\big) - a\Big) \\ &= -B\big(\pi\big(\mathrm{ad}(y)x\big), \pi(x) - a\big) = -B\big(\mathrm{ad}(y)x, \pi(x) - a\big).\end{aligned}$$

Clearly $\pi(x) - a \in \mathfrak{a}$ and so $\varphi_x = a - \pi(x)$. The corresponding gradient flow is then (compare [4, Example 2])

$$\frac{dx}{dt} = \big[x, \big[x, a - \pi(x)\big]\big].$$

Thus $x(\infty) = \lim_{t\to\infty} x(t) \in \Omega$ exists and $[x(\infty), a - \pi(x(\infty))] = 0$.

(1) Now when $G = GL(n, \mathbb{R})$, $K = O(n)$, $B(x, y) = \operatorname{tr} xy$, $\theta : x + y \mapsto x - y$, $x \in \mathfrak{k}$ and $y \in \mathfrak{p}$, where $\mathfrak{p}$ is the set of real $n \times n$ symmetric matrices, $\mathfrak{k}$ is the set of real $n \times n$ skew symmetric matrices, $\mathfrak{a} \subset \mathfrak{p}$ is the set of real diagonal matrices, π means taking diagonal component of elements in $\mathfrak{p}$. Thus the gradient flow becomes [5, p. 211]

$$\frac{dx}{dt} = \big[x, \big[x, a - \mathrm{diag}(x)\big]\big],$$

where $x(t)$ is a real $n \times n$ symmetric matrix. Along this flow

$$\frac{d\varphi}{dt} = -B\big(\big[x, a - \mathrm{diag}(x)\big], \big[x, a - \mathrm{diag}(x)\big]\big).$$

The results remain the same for $G = GL(n, \mathbb{C})$, that is, the Hermitian case.

(2) When $G = O(p, q)$, the gradient flow takes the form

$$\begin{aligned}\frac{dX}{dt} &= \big[A - \mathrm{diag}(X)\big]X^{\mathrm{T}}X - X\big[A - \mathrm{diag}(X)\big]^{\mathrm{T}}X \\ &\quad + XX^{\mathrm{T}}\big[A - \mathrm{diag}(X)\big] - X\big[A - \mathrm{diag}(X)\big]^{\mathrm{T}}X,\end{aligned}$$

where

$$x(t) = \begin{bmatrix} 0 & X(t) \\ X(t)^{\mathrm{T}} & 0 \end{bmatrix}, \qquad a = \begin{bmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{bmatrix},$$

and A is a $p \times q$ “diagonal” matrix.

The following extends a result in [6, p. 211].

Theorem 2.9. *The gradient flow of $\Phi : K \to \mathbb{R}$ defined by*

$$\Phi(k) = -\frac{1}{2}B\big(\pi\big(\mathrm{Ad}\big(k^{-1}\big)x_0\big) - a, \pi\big(\mathrm{Ad}\big(k^{-1}\big)x_0\big) - a\big),$$

where $x_0 \in \Omega$ and $a \in \mathfrak{a}$, with respect to the left-invariant Riemannian metric on $\mathfrak{k}$ is

$$\frac{dk}{dt} = (dL)_{k(t)}\big[\mathrm{Ad}\big(k(t)^{-1}\big)x_0, a - \pi\big(\mathrm{Ad}\big(k(t)^{-1}\big)x_0\big)\big].$$

The projection of the gradient flow onto the adjoint orbit Ω obtained by setting $x(t) = \mathrm{Ad}(k(t)^{-1})x_0$ is the double bracket equation

$$\frac{dx}{dt} = \big[x, \big[x, a - \pi(x)\big]\big].$$

Proof. As before, let $v = (dL)_k r$ where $r \in \mathfrak{k}$. Now

$$\begin{aligned}
(d\Phi)_k v &= \frac{d}{dt}\bigg|_{t=0} \Phi\big(ke^{tr}\big) \\
&= -\frac{1}{2}\frac{d}{dt}\bigg|_{t=0} B\Big(\pi\big(\mathrm{Ad}\big(ke^{tr}\big)^{-1}x_0\big) - a, \pi\big(\mathrm{Ad}\big(ke^{tr}\big)^{-1}x_0\big) - a\Big) \\
&= -B\big(\pi\big(\big[\mathrm{Ad}\big(k^{-1}\big)x_0, r\big]\big), \pi\big(\mathrm{Ad}\big(k^{-1}\big)x_0\big) - a\big) \\
&= -B\big(\big[\mathrm{Ad}\big(k^{-1}\big)x_0, r\big], \pi\big(\mathrm{Ad}\big(k^{-1}\big)x_0\big) - a\big) \quad \big(\text{since } \pi\big(\mathrm{Ad}\big(k^{-1}\big)x_0\big) - a \in \mathfrak{a}\big) \\
&= -B\big(\big[\mathrm{Ad}\big(k^{-1}\big)x_0, a - \pi\big(\mathrm{Ad}\big(k^{-1}\big)x_0\big)\big], r\big) \\
&= \big\langle (dL)_k\big[\mathrm{Ad}\big(k^{-1}\big)x_0, a - \pi\big(\mathrm{Ad}\big(k^{-1}\big)x_0\big)\big], v\big\rangle.
\end{aligned}$$

Thus the gradient flow takes the desired form. Now [6, p. 211]

$$\frac{dx}{dt} = \big[x, (dL)_k^{-1}(dL)_k\big[x, a - \pi(x)\big]\big] = \big[x, \big[x, a - \pi(x)\big]\big]. \qquad \square$$

Remark. The maximal abelian subalgebra $\mathfrak{a}$ in $\mathfrak{p}$ in Examples 2.7 and 2.8 can be replaced by any subspace S of $\mathfrak{p}$ and $\pi : \mathfrak{p} \to S$ is the orthogonal projection and the results still hold.

Example 2.10. With the same notations in Lemma 2.5, let $\varphi : \Omega \to \mathbb{R}$ be given by

$$\varphi(x) = -\frac{1}{2}B\big(x - \pi(x), x - \pi(x)\big), \quad x \in \Omega.$$

For all $y \in \mathfrak{k}_x^{\perp}$, we have

$$\begin{aligned}
(d\varphi)_x\big(\mathrm{ad}(y)x\big) &= -\frac{1}{2}\frac{d}{dt}\bigg|_{t=0} B\Big(\mathrm{Ad}\big(e^{ty}\big)x - \pi\big(\mathrm{Ad}\big(e^{ty}\big)x\big), \mathrm{Ad}\big(e^{ty}\big)x - \pi\big(\mathrm{Ad}\big(e^{ty}\big)x\big)\Big) \\
&= -B\big(\mathrm{ad}(y)x - \pi\big(\mathrm{ad}(y)x\big), x - \pi(x)\big) \\
&= -B\big(\mathrm{ad}(y)x, x - \pi(x)\big) + B\big(\pi\big(\mathrm{ad}(y)x\big), x - \pi(x)\big) \\
&= -B\big(\mathrm{ad}(y)x, x - \pi(x)\big) + B\big(\pi\big(\mathrm{ad}(y)x\big), \pi\big(x - \pi(x)\big)\big) \\
&= -B\big(\mathrm{ad}(y)x, x - \pi(x)\big) \quad \big(\text{since } \pi\big(x - \pi(x)\big) = \pi(x) - \pi(x) = 0\big) \\
&= -B\big(\mathrm{ad}(y)x, x\big) + B\big(\mathrm{ad}(y)x, \pi(x)\big)
\end{aligned}$$

$$
\begin{aligned}
&= -B\big([y,x],x\big) + B\big(\operatorname{ad}(y)x, \pi(x)\big)\\
&= B\big(y,[x,x]\big) + B\big(\operatorname{ad}(y)x, \pi(x)\big)\\
&= B\big(\operatorname{ad}(y)x, \pi(x)\big).
\end{aligned}
$$

So $(d\varphi)_x(\operatorname{ad}(y)x) = B(\operatorname{ad}(y)x, \pi(x))$. Thus $\varphi_x = \pi(x)$. The corresponding gradient flow is (compare [4, Example 2])

$$\frac{dx}{dt} = \big[x, \big[x, \pi(x)\big]\big].$$

Thus $x(\infty) = \lim_{t\to\infty} x(t) \in \Omega$ exists and $[x(\infty), \pi(x(\infty))] = 0$.

The following extends [4, Example 2] and [7, p. 1057].

Theorem 2.11. *The gradient flow of $\Phi : K \to \mathbb{R}$ defined by*

$$\Phi(k) = -\frac{1}{2} B\big(\operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big), \operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big),$$

where $x_0 \in \Omega$ and $a \in \mathfrak{a}$, with respect to the left-invariant Riemannian metric on $\mathfrak{k}$ is

$$\frac{dk}{dt} = (dL)_k\big[\operatorname{Ad}\big(k^{-1}\big)x_0, \operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big].$$

The projection of the gradient flow onto the adjoint orbit Ω obtained by setting $x(t) = \operatorname{Ad}(k(t)^{-1})x_0$ is the double bracket equation

$$\frac{dx}{dt} = \big[x, \big[x, \pi(x)\big]\big].$$

Proof. Let $v = (dL)_k r$ where $r \in \mathfrak{k}$. Now

$$
\begin{aligned}
(d\Phi)_k v &= \frac{d}{dt}\bigg|_{t=0} \Phi\big(ke^{tr}\big)\\
&= -\frac{1}{2}\frac{d}{dt}\bigg|_{t=0} B\big(\operatorname{Ad}\big(ke^{tr}\big)^{-1}x_0 - \pi\big(\operatorname{Ad}\big(ke^{tr}\big)^{-1}x_0\big), \operatorname{Ad}\big(ke^{tr}\big)^{-1}x_0 - \pi\big(\operatorname{Ad}\big(ke^{tr}\big)^{-1}x_0\big)\big)\\
&= -B\big(\big[\operatorname{Ad}\big(k^{-1}\big)x_0, r\big] - \pi\big(\big[\operatorname{Ad}\big(k^{-1}\big)x_0, r\big]\big), \operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big)\\
&= -B\big(\big[\operatorname{Ad}\big(k^{-1}\big)x_0, r\big], \operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big)\\
&\quad + B\big(\pi\big(\big[\operatorname{Ad}\big(k^{-1}\big)x_0, r\big]\big), \operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big)\\
&= B\big(\big[\operatorname{Ad}\big(k^{-1}\big)x_0, \operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big], r\big)\\
&\quad + B\big(\pi\big(\big[\operatorname{Ad}\big(k^{-1}\big)x_0, r\big]\big), \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big)\big)\\
&= B\big(\big[\operatorname{Ad}\big(k^{-1}\big)x_0, \operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big], r\big)\\
&\quad \big(\text{since } \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big) = 0\big)\\
&= B\big((dL)_k\big[\operatorname{Ad}\big(k^{-1}\big)x_0, \operatorname{Ad}\big(k^{-1}\big)x_0 - \pi\big(\operatorname{Ad}\big(k^{-1}\big)x_0\big)\big], v\big).
\end{aligned}
$$

Thus the gradient flow takes the required form. Now

$$\frac{dx}{dt} = \left[x, (dL)_k^{-1}(dL)_k\left[x, x - \pi(x)\right]\right] = \left[x, \left[x, x - \pi(x)\right]\right]. \qquad \square$$

Example 2.12.

(1) When $G = GL(n, \mathbb{R})$, the gradient flow becomes

$$\frac{dx}{dt} = \left[x, \left[x, x - \operatorname{diag}(x)\right]\right],$$

where $x(t)$ is a real $n \times n$ symmetric matrix. The results remain the same for $G = GL(n, \mathbb{C})$, that is, the Hermitian case.

(2) When $G = O(p, q)$, the gradient flow takes the form

$$\begin{aligned}\frac{dX}{dt} &= \left[X - \operatorname{diag}(X)\right]X^{\mathrm{T}}X - X\left[X - \operatorname{diag}(X)\right]^{\mathrm{T}}X \\ &\quad + XX^{\mathrm{T}}\left[X - \operatorname{diag}(X)\right] - X\left[X - \operatorname{diag}(X)\right]^{\mathrm{T}}X,\end{aligned}$$

where

$$x(t) = \begin{bmatrix} 0 & X(t) \\ X(t)^{\mathrm{T}} & 0 \end{bmatrix},$$

and $X(t)$ is a real $p \times q$ matrix. The result is similar for $U(p, q)$.

3. Global extrema

We now discuss the global extrema of the three optimization problems. We may consider the problems in a more general setting. Let $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ be the Cartan decomposition of a reductive Lie algebra $\mathfrak{g}$. Of course $\mathfrak{g}_1 := [\mathfrak{g}, \mathfrak{g}]$ is semisimple. Let K be the analytic subgroup of G_1 for $\mathfrak{k}$. Thus $\operatorname{Ad}(K)$ is a maximal compact subgroup of $\operatorname{Ad}(G_1)$ though K may not be maximal compact in G_1. It is known that $\operatorname{Ad}(K)$ leaves $\mathfrak{p}$ invariant. Let $\mathfrak{a}$ be a maximal abelian subalgebra in $\mathfrak{p}$ with a fixed (closed) fundamental Weyl chamber $\mathfrak{a}_+$ and let $\pi : \mathfrak{p} \to \mathfrak{a}$ be the orthogonal projection with respect to the Killing form B. Let Ω be the orbit of an element in $\mathfrak{p}$ under the adjoint action of $\operatorname{Ad}(K)$.

(1) The global maximum and minimum of $\varphi(x) = B(x, z)$ are obtained in [14, Theorem 2.1] as well as the maximizers and minimizers. Let x_0 denote the unique element of the singleton set $\Omega \cap \mathfrak{a}_+$ where $\mathfrak{a}_+$ is a closed fundamental Weyl chamber of a maximal abelian subalgebra in $\mathfrak{p}$. The minimum of φ is $B(x_0, \omega_0 z_0)$ and the maximum is $B(x_0, z_0)$ where $\operatorname{Ad}(K) \cap \mathfrak{a}_+ = \{z_0\}$ and ω_0 is the longest element of the Weyl group W of $(\mathfrak{g}, \mathfrak{a})$. The maximum (minimum) is achieved at $\varphi(x)$ if and only if there is $k \in K$ such that both $\operatorname{Ad}(k)x$ and $\operatorname{Ad}(k)z$ are in $\mathfrak{a}_+$ ($\operatorname{Ad}(k)x \in \mathfrak{a}_+$ but $\operatorname{Ad}(k)z \in -\mathfrak{a}_+$).

(2) By Kostant's convexity theorem [13] or [10, Theorem 3.6], $\pi(\Omega) = \operatorname{conv} Wx_0$ where $\Omega \cap \mathfrak{a}_+ = \{x_0\}$, W is the Weyl group of $(\mathfrak{g}, \mathfrak{a})$, conv denotes the convex hull of the orbit Wx_0, and $\pi : \mathfrak{p} \to \mathfrak{a}$ is the orthogonal projection with respect to the Killing form. The global minimum of φ is $-\frac{1}{2}\max_{x \in \operatorname{conv} Wx_0} B(a - x, a - x)$. The global maximum is $-\frac{1}{2}\min_{x \in \operatorname{conv} Wx_0} B(a - x, a - x)$; see

[10, Corollary 2.14] for the minimizer and the algorithm for finding the global maximizer. Both are related to the Weyl group W. Now if a is in the convex hull of the orbit of $\Omega \cap \mathfrak{a}_+$ under the action of W, then of course the global maximum of φ is zero.

(3) On the last problem we have the following theorem.

Theorem 3.1. *Let $\varphi : \Omega \to \mathbb{R}$ be defined by $\varphi(x) = -\frac{1}{2}B(x - \pi(x), x - \pi(x))$ where $\pi : \mathfrak{p} \to \mathfrak{a}$ is the orthogonal projection onto a maximal abelian subalgebra $\mathfrak{a}$ in $\mathfrak{p}$. Then φ attains its global minimum 0 at those $x \in \mathfrak{a}$ and attains its global maximum at x such that $c(\Omega) := \frac{1}{|W|}\sum_{w \in W} wy \in \mathfrak{a}$, where $y \in \Omega \cap \mathfrak{a}$ and W is the Weyl group of $(\mathfrak{g}, \mathfrak{a})$. The element $c(\Omega) \in \mathfrak{a}$ is independent of the choice of $y \in \Omega \cap \mathfrak{a}$.*

Proof. Let $y \in \Omega \cap \mathfrak{a}$ [10, Lemma 3.4]. Notice that

$$B\big(x - \pi(x), x - \pi(x)\big) = B(x, x) + B\big(\pi(x), \pi(x)\big) - 2B\big(x, \pi(x)\big) = B(x, x) - B\big(x, \pi(x)\big),$$

since $B(x, \pi(x)) = B(\pi(x), \pi(x))$. Now $B(x, x)$ is constant on Ω and thus finding the global maximum and global minimum amount to finding the global minimum and global maximum of $B(\pi(x), \pi(x))$ respectively. Of course, the global maximum of $B(\pi(x), \pi(x))$ is $B(x, x)$ attainable at Wy. Hence $\min_{x \in \Omega} \varphi(x) = 0$.

Now $\pi(x)$ is in the convex hull of the orbit of y, Wy, under the action of the Weyl group W [10], by Kostant's convexity theorem [13] or [10, Theorem 3.6], that is, $\sum_{w \in W} \alpha_w wy$ where $\sum_{w \in W} \alpha_w = 1$ and $\alpha_w \geqslant 0$ for all $w \in W$, and the norm function $\psi(z) := B(z, z)^{1/2}$ is evidently convex and Ad-invariant on $\mathfrak{p}$ and thus convex and W-invariant on $\mathfrak{a}$ (that is, $\psi(wz) = \psi(z)$ for all $w \in W$, $z \in \mathfrak{a}$). Consider the element (which can be viewed as the center of the convex hull of $\Omega \cap \mathfrak{a}$)

$$c(\Omega) := \frac{1}{|W|}\sum_{w \in W} wy$$

which is independent of the choice of $y \in \Omega \cap \mathfrak{a}$ since they are conjugate under the action of the Weyl group, that is, $wy = y'$ for some $w \in W$ if $y, y' \in \Omega \cap \mathfrak{a}$. Now $c(\Omega)$ is in the convex hull of $W(\pi(x))$ since

$$\frac{1}{|W|}\sum_{w \in W} w\pi(x) = \frac{1}{|W|}\sum_{w' \in W} w' \sum_{w \in W} \alpha_w wy = \frac{1}{|W|}\sum_{w \in W} wy = c(\Omega).$$

So $B(\pi(x), \pi(x)) \geqslant B(c, c)$ which is the global minimum. $\square$

Remark.

(1) The operator in the above proof $c := \frac{1}{|W|}\sum_{w \in W} w : \mathfrak{a} \to \mathfrak{a}$ is idempotent, that is, $c^2 = c$.

(2) When $\mathfrak{g}$ is semisimple, then $c(\Omega)$ is zero since the Weyl group is then a direct sum of irreducible Weyl groups and thus is essential, that is, the set of fixed points of $\mathfrak{a}$ under the action of W is trivial. However, $c(\Omega)$ is obviously fixed by W. For example, $G = SL(n, \mathbb{R})$, $\mathfrak{p}$ is the set of traceless real $n \times n$ symmetric matrices and the Weyl group is the symmetric group which is essential to $\mathfrak{a}$ (the set of $n \times n$ real diagonal matrices).

Example 3.2. When $G = GL(n, \mathbb{R})$, those diagonal matrices cospectral with a given real $n \times n$ symmetric matrix x are global minimizers. The matrix

$$\frac{1}{n}(\operatorname{tr} x)I = \sum_{w \in S_n} \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

is a global minimizer since the Weyl group is the symmetric group S_n, where $x \in \Omega$ and $\operatorname{diag}(\lambda_1, \dots, \lambda_n) \in \Omega \cap \mathfrak{a}$, that is, $\lambda_1, \dots, \lambda_n$ are the eigenvalues of x. The minimum of φ is

$$-\frac{1}{2}\left\| x - \frac{1}{n}(\operatorname{tr} x)I \right\|^2 = -\frac{1}{2}\bigl(\operatorname{tr} x^2 - (\operatorname{tr} x)^2\bigr),$$

where $\|\cdot\|$ is the Frobenius norm; see [7, Theorem 4.2].

Applying the result on $O(p,q)$, we can derive the corresponding result on the set of real $p \times q$ matrices possessing prescribed singular values. The global maximizers are those s in the singular value decomposition $x = usv$, $u \in O(p)$ and $v \in O(q)$, that is, the "diagonal" matrices with singular values of x on the diagonal. The zero matrix is a global minimizer which gives $-\frac{1}{2}\|x\|^2 = -\frac{1}{2}\operatorname{tr} xx^{\mathrm{T}}$ as the minimum; see [7, Theorem 5.2].

4. Remark

Applying the results on various real simple Lie groups, for example, classical groups, different gradient flows will be obtained and the interested readers may work out the details. These include the set of complex skew symmetric matrices and the set complex symmetric matrices with prescribed singular values, etc.

The compact Lie group K in the consideration of [2] which is not connected may not be covered by our approach. For example, consider $K = O(2n)$ [12, p. 385]. It has two components $O^+(2n) := SO(2n)$ and $O^-(2n)$ (determinant -1 orthogonal matrices). Its Lie algebra $\mathfrak{o}(2n)$ is the set of all $2n \times 2n$ real skew symmetric matrices. The adjoint action of $O^+(2n)$ preserves the Pfaffian but the adjoint action of $O^-(2n)$ changes the sign of the Pfaffian since for any $k \in O(2n)$,

$$\operatorname{Pf}\bigl(kxk^{-1}\bigr) = \operatorname{Pf}\bigl(kxk^{\mathrm{T}}\bigr) = \det(k)\operatorname{Pf}(x).$$

Let $x_0 \in \mathfrak{o}(2n)$. If x_0 is nonsingular, the adjoint orbit Ω has two components Ω_+ and Ω_- corresponding to $O^+(2n)$ and $O^-(2n)$ respectively. The gradient flow remains in the same component Ω_+ or Ω_- since the flow is continuous. The image of Ω may be disconnected under the map

$$\varphi : \Omega \to \mathbb{R}, \quad x \mapsto -B(x,z).$$

Now $-B(x,z) = -\operatorname{tr} xz$. When $n = 1$, that is,

$$\Omega = \left\{ \begin{bmatrix} 0 & y \\ -y & 0 \end{bmatrix} : \ y = \pm\alpha \right\}$$

and if

$$z = \begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix},$$

then $\varphi(\Omega) = \{2\alpha\beta, -2\alpha\beta\}$, a set of two points.

The Hessian of Φ with respect to the first problem has been fully studied in [8]. It certainly has merits in numerical analysis [7] in which the stability of the equilibria (critical points) is a concern, when the underlying reductive groups are $GL(n, \mathbb{R})$ and $O(p, q)$.

References

[1] R. Bott, On torsion in Lie groups, Proc. Nat. Acad. Sci. USA 40 (1954) 586–588.
[2] A.M. Bloch, R.W. Brockett, T. Ratiu, Completely integrable gradient flows, Comm. Math. Phys. 147 (1992) 57–74.
[3] R.W. Brockett, Dynamical system that sort lists, diagonalize matrices and solving least squares programming problems, Linear Algebra Appl. 146 (1991) 79–91.
[4] R.W. Brockett, Differential geometry and the design of gradient algorithms, Proc. Symp. Pure Math. 54 (I) (1993) 69–92.
[5] M.T. Chu, Constructing a Hermitian matrix from its diagonal entries and eigenvalues, SIAM J. Matrix Anal. Appl. 16 (1995) 207–217.
[6] M.T. Chu, Inverse eigenvalue problems, SIAM Rev. 40 (1998) 1–39.
[7] M.T. Chu, K. Driessel, The projected gradient method for least squares matrix approximations with spectral constraints, SIAM J. Numer. Anal. 27 (1990) 1050–1060.
[8] J.J. Duistermaat, J.A.C. Kolk, V.S. Varadarajan, Functions, flows and oscillatory integrals on flag manifolds and conjugacy classes in real semisimple Lie groups, Compositio Math. 49 (1983) 309–398.
[9] S. Helgason, Differential Geometry, Lie Groups and Symmetric Spaces, Academic Press, San Diego, CA, 1978.
[10] R.R. Holmes, T.Y. Tam, Distance to the convex hull of an orbit under the action of a compact Lie group, J. Austral. Math. Soc. Ser. A 66 (1999) 331–357.
[11] G.A. Hunt, A theorem of Élie Cartan, Proc. Amer. Math. Soc. 7 (1956) 307–308.
[12] A.W. Knapp, Lie Groups Beyond an Introduction, Birkhäuser, Boston, 1996.
[13] B. Kostant, On convexity, the Weyl group and Iwasawa decomposition, Ann. Sci. Ecole Norm. Sup. (4) 6 (1973) 413–460.
[14] T.Y. Tam, An extension of a result of Lewis, Electronic J. Linear Algebra 5 (1999) 1–10.
[15] V.S. Varadarajan, The method of stationary phase and applications to geometry and analysis on Lie groups, in: B. Orsted, H. Schlichtkrull (Eds.), Algebraic and Analytic Methods in Representation Theory, Perspectives in Mathematics, vol. 17, Academic Press, 1997, pp. 167–242.
[16] D.A. Vogan Jr., The method of stationary phase and applications to geometry and analysis on Lie groups, in: B. Orsted, H. Schlichtkrull (Eds.), Algebraic and Analytic Methods in Representation Theory, Perspectives in Mathematics, vol. 17, Academic Press, 1997, pp. 243–339.